

IN THE CLAIMS:

The following is a complete listing of the claims, and replaces all earlier listings and all earlier versions.

1. (Currently Amended) A document segmentation apparatus comprising:

A¹ table analyzing means for generating cell position data indicating a positional relationship between cells and cell vectors representing characteristics of the cells, by analyzing a table pinched between a start tag and an end tag in a document to be processed;

table type judging means for judging a table type with reference to the cell position data and the cell vectors generated by said table analyzing means;

first segment generating means for generating a ~~segment from~~ plurality of segments each of which is pinched between the start tag and the end tag by dividing the table when with a first method in a case in which the table type is a ~~table for showing a table~~ list type; and

second segment generating means for generating a ~~segment from~~ plurality of segments each of which is pinched between the start tag and the end tag by dividing the table when with a second method in a case in which the table type is a ~~table for layout~~ type.

2. (Currently Amended) A document segmentation apparatus according to claim 1, wherein said first segment generating means comprise[[:]]:

cut direction determination means for determining a cut direction of the table by judging whether the data is expressed in a column or a row in the table on the basis of the cell position data and the cell vectors; and

table segment generating means for generating a table segment by dividing the table on the basis of the table type and the cut direction.

3. (Original) A document segmentation apparatus according to claim 2, wherein said second segment generating means generate the table itself as the segment.

4. (Currently Amended) A document segmentation apparatus according to claim 1, wherein said second segment generating means comprise[[]];

cell cluster generating means for generating cell cluster information by clustering the cells in the table; and

layout segment generating means for generating segment by connecting the cells in the table with reference to the cell position data and the cell cluster information.

5. (Original) A document segmentation apparatus according to claim 4, wherein said first segment generating means generate the table itself as the segment.

6. (Original) A document segmentation apparatus according to claim 4, wherein said second segment generating means generate the table itself as the segment.

7. (Currently Amended) A document segmentation apparatus according to claim 1, further comprising:

_____ normal segment generating means for dividing the document into a segment which corresponds to one table[[]],

[[and]] wherein the table generated as one segment by said normal segment generating means is to be processed by said table analyzing means.

A (8. (Original) A document segmentation apparatus according to claim 1, wherein said table analyzing means further generate cell data of the analyzed table and said table and said table type judging means judge the table type with reference to the cell data.

9. (Original) A document segmentation apparatus according to claim 8, wherein said table type judging means comprise similarity judging means for judging the table type on the basis of similarity between the cell data positioned at particular positions with reference to the cell position data and the cell data generated by said table analyzing means.

10. (Original) A document segmentation apparatus according to claim 8, wherein said table type judging means comprise partial character line extracting means for extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated by said table analyzing means,

and character line comparing means for comparing the extracted partial character lines to judge the table type.

11. (Original) A document segmentation apparatus according to claim 8, wherein said table type judging means comprise partial character line extracting means for extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated by said table analyzing means, and similarity judging means for judging the table type on the basis of similarity between the extracted partial character lines.

12. (Original) A document segmentation apparatus according to claim 8, wherein said table type judging means comprise syntax judging means for judging the table type with reference to the cell position data, the cell vectors and the cell data generated by said table analyzing means, and similarity judging means for judging the table type on the basis of similarity between the cell data positioned at particular positions with reference to the cell position data and the cell data generated by said table analyzing means.

13. (Original) A document segmentation apparatus according to claim 8, wherein said table type judging means comprise syntax judging means for judging the table type with reference to the cell position data, the cell vectors and the cell data generated by said table analyzing means, partial character line extracting means for extracting partial character lines from the cell data positioned at a particular position with reference to the

cell position data and the cell data generated by said table analyzing means, and, character line comparing means for comparing the extracted partial character lines to judge the table type.

41
14. (Original) A document segmentation apparatus according to claim 8, wherein said table type judging means comprise syntax judging means for judging the table type with reference to the cell position data, the cell vectors and the cell data generated by said table analyzing means, partial character line extracting means for extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated by said table analyzing means, and similarity judging means for judging the table type on the basis of similarity between the extracted partial character lines.

15. (Currently Amended) A document segmentation apparatus according to claim 1, further comprising:

————— table reforming means for reforming the table so that the number of cells in each column and each row becomes the same, by analyzing the table to be processed[[;]],

[[and]] wherein said table analyzing means analyze the reformed table.

16. (Original) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise supplementary data removing means for removing data added to the table from the table data.

17. (Original) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise multi-row/multi-column processing means for reforming the table regularly by analyzing the structure of the table data.

18. (Original) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise composite table processing means for reforming the table by analyzing regularity of information description constituting the table.

19. (Currently Amended) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise[[:]]:

supplementary data removing means for removing data added to the table from the table data; and

multi-row/multi-column processing means for reforming the table regularly by analyzing the structure of the table data.

20. (Currently Amended) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise[[:]]:

supplementary data removing means for removing data added to the table from the table data; and

composite table processing means for reforming the table by analyzing regularity of information description constituting the table.

21. (Currently Amended) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise[[:]]:

multi-row/multi-column processing means for reforming the table regularly by analyzing the structure of the table data; and

composite table processing means for reforming the table by analyzing regularity of information description constituting the table.

22. (Original) A document segmentation apparatus according to claim 15, wherein said table reforming means comprise:

supplementary data removing means for removing data added to the table from the table data;

multi-row/multi-column processing means for reforming the table regularly by analyzing the structure of the table data; and

composite table processing means for reforming the table by analyzing regularity of information description constituting the table.

23. (Currently Amended) A document segmentation method comprising:
a table analyzing step of [[for]] generating cell position data indicating a positional relationship between cells and cell vectors representing characteristics of the cells, by analyzing a table pinched between a start tag and an end tag in a document to be processed;

a table type judging step_{of} [[for]] judging a table type with reference to the cell position data and the cell vectors generated [[by]] in said table analyzing step;

a first segment generating step_{of} [[for]] generating a ~~segment from~~ plurality of segments each of which is pinched between the start tag and the end tag by dividing the table ~~when~~ with a first method in a case in which the table type is a ~~table~~ describing a table list format; and

A (a second segment generating step_{of} [[for]] generating a ~~segment from~~ plurality of segments each of which is pinched between the start tag and the end tag by dividing the table ~~when~~ with a second method in a case in which the table type is a ~~table for~~ layout type.

24. (Currently Amended) A document segmentation method according to claim 23, wherein said first segment generating step comprises:

a cut direction determination step_{of} [[for]] determining a cut direction of the table by judging whether the data is expressed in a column or a row in the table on the basis of the cell position data and the cell vectors; and

a table segment generating step_{of} [[for]] generating a table segment by dividing the table on the basis of the table type and the cut direction.

25. (Currently Amended) A document segmentation method according to claim 24, wherein said second segment generating step ~~generates~~ includes generating the table itself as the segment.

26. (Currently Amended) A document segmentation method according to claim 23, wherein said second segment generating step comprises[[:]]:

a cell cluster generating step_{of} [[for]] generating cell cluster information by clustering the cells in the table; and

a layout segment generating step_{of} [[for]] generating segment by connecting the cells in the table with reference to the cell position data and the cell cluster information.

A (

27. (Currently Amended) A document segmentation method according to claim 26, wherein said first segment generating step ~~generates~~ includes generating the table itself as the segment.

28. (Currently Amended) A document segmentation method according to claim 26, wherein said second segment generating step ~~generates~~ includes generating the table itself as the segment.

29. (Currently Amended) A document segmentation method according to claim 23, further comprising

————— a normal segment generating step_{of} [[for]] dividing the document into a segment which corresponds to one table[[:]],

[[and]] wherein the table generated as one segment [[by]] in said normal segment generating step is to be processed [[by]] in said table analyzing step.

30. (Currently Amended) A document segmentation method according to claim 23, wherein said table analyzing step further ~~generates~~ includes generating cell data of the analyzed table and said table type judging step ~~judges~~ includes judging the table type with reference to the cell data.

31. (Currently Amended) A document segmentation method according to claim 30, wherein said table type judging step comprises a similarity judging step, of ~~[[for]]~~ judging the table type on the basis of similarity between the cell data positioned at particular positions with reference to the cell position data and the cell data generated ~~[[by]]~~ in said table analyzing step.

32. (Currently Amended) A document segmentation method according to claim 30, wherein said table type judging step comprises a partial character line extracting step, of ~~[[for]]~~ extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated ~~[[by]]~~ in said table analyzing step, and a character line comparing step, of ~~[[for]]~~ comparing the extracted partial character lines to judge the table type.

33. (Currently Amended) A document segmentation method according to claim 30, wherein said table type judging step comprises a partial character line extracting ~~means for~~ step, of extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated

[[by]] in said table analyzing step, and a similarity judging step, of [[for]] judging the table type on the basis of similarity between the extracted partial character lines.

34. (Currently Amended) A document segmentation method according to claim 30, wherein said table type judging step comprises a syntax judging step, of [[for]] judging the table type with reference to the cell position data, the cell vectors and the cell data generated [[by]] in said table analyzing step, and a similarity judging step, of [[for]] judging the table type on the basis of similarity between the cell data positioned at particular positions with reference to the cell position data and the cell data generated [[by]] in said table analyzing step.

35. (Currently Amended) A document segmentation method according to claim 30, wherein said table type judging step comprises a syntax judging step, of [[for]] judging the table type with reference to the cell position data, the cell vectors and the cell data generated [[by]] in said table analyzing step, a partial character line extracting step, of [[for]] extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated [[by]] in said table analyzing step, and a character line comparing step, of [[for]] comparing the extracted partial character lines to judge the table type.

36. (Currently Amended) A document segmentation method according to claim 30, wherein said table type judging step comprises a syntax judging step, of [[for]]

judging the table type with reference to the cell position data, the cell vectors and the cell data generated [[by]] in said table analyzing step, a partial character line extracting step, of [[for]] extracting partial character lines from the cell data positioned at a particular position with reference to the cell position data and the cell data generated [[by]] in said table analyzing step, and a similarity judging ~~means for~~ step, of judging the table type on the basis of similarity between the extracted partial character lines.

37. (Currently Amended) A document segmentation method according to claim 23, further comprising—

———— a table reforming step, of [[for]] reforming the table so that the number of cells in each column and each row becomes the same, by analyzing the table to be processed~~[[;]]~~,

[[and]] wherein said table analyzing step ~~analyzes~~ includes analyzing the reformed table.

38. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises a supplementary data removing step, of [[for]] removing data added to the table from the table data.

39. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises a multi-row/multi-column

processing step, of [[for]] reforming the table regularly by analyzing the structure of the table data.

40. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises a composite table processing step, of [[for]] reforming the table by analyzing regularity of information description constituting the table.

41. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises[[;]]:

a supplementary data removing step, of [[for]] removing data added to the table from the table data; and

a multi-row/multi-column processing step, of [[for]] reforming the table regularly by analyzing the structure of the table data.

42. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises[[;]]:

a supplementary data removing step, of [[for]] removing data added to the table from the table data; and

a composite table processing step, of [[for]] reforming the table by analyzing regularity of information description constituting the table.

43. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises[[:]]:

a multi-row/multi-column processing step_{of} [[for]] reforming the table regularly by analyzing the structure of the table data; and

a composite table processing step_{of} [[for]] reforming the table by analyzing regularity of information description constituting the table.

44. (Currently Amended) A document segmentation method according to claim 37, wherein said table reforming step comprises[[:]]:

a supplementary data removing step_{of} [[for]] removing data added to the table from the table data;

a multi-row/multi-column processing step_{of} [[for]] reforming the table regularly by analyzing the structure of the table data; and

a composite table processing step_{of} [[for]] reforming the table by analyzing regularity of information description constituting the table.

45. (Currently Amended) A computer-readable storage medium storing a document segmentation program for controlling a computer to perform document segmentation, said program comprising codes for causing the computer to perform:

a table analyzing step_{of} [[for]] generating cell position data indicating a positional relationship between cells and cell vectors representing characteristics of the

cells, by analyzing a table pinched between a start tag and an end tag in a document to be processed;

a table type judging step, of [[for]] judging a table type with reference to the cell position data and the cell vectors generated [[by]] in said table analyzing step;

A (a first segment generating step, of [[for]] generating a ~~segment from~~ plurality of segments each of which is pinched between the start tag and the end tag by dividing the table when with a first method in a case in which the table type is a table describing a table list format; and

a second segment generating step, of [[for]] generating a ~~segment from~~ plurality of segments each of which is pinched between the start tag and the end tag by dividing the table when with a second method in a case in which the table type is a ~~table for~~ layout type.
